Congratulations! You passed!

Grade received 80% | Latest Submission Grade 80% | To pass 80% or higher

Go to next item

1. You are building a 3-class object classification and localization algorithm. The classes are: pedestrian (c=1), car (c=2), motorcycle (c=3). What should y be for the image below? Remember that "?" means "don't care", which means that the neural network loss function won't care what the neural network gives for that component of the output. Recall $y = [p_c, b_x, b_y, b_h, b_w, c_1, c_2, c_3]$. 1 / 1 point

https://www.pexels.com/es-es/foto/mujer-vestida-con-falda-azul-y-blanca-caminando-cerca-de-la-hierba-verde-durante-el-dia-144474/

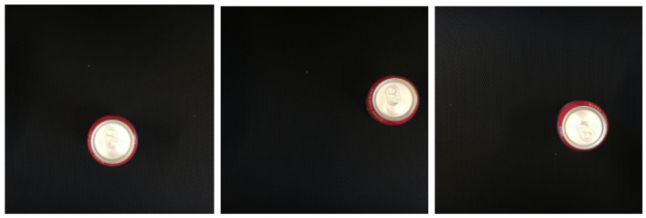
- $y = [1, 0.66, 0.5, 0.75, 0.16, 0, 0, 0]$
- $y = [1, ?, ?, ?, ?, 1, ?, ?]$
- $y = [1, 0.66, 0.5, 0.75, 0.16, 1, 0, 0]$

Loading [MathJax]/jax/output/CommonHTML/jax.js

Expand

Correct
Correct. $p_c = 1$ since there is a pedestrian in the picture. We can see that b_x, b_y as percentages of the image are approximately correct as well b_h, b_w, and the value of $c_1 = 1$ for a pedestrian.

2. You are working on a factory automation task. Your system will see a can of soft-drink coming down a conveyor belt, and you want it to take a picture and decide whether (i) there is a soft-drink can in the image, and if so (ii) its bounding box. Since the soft-drink can is round, the bounding box is always square, and the soft drink can always appear the same size in the image. There is at most one soft drink can in each image. Here're some typical images in your training set: 1 / 1 point



To solve this task it is necessary to divide the task into two: 1. Construct a system to detect if a can is present or not. 2. Construct a system that calculates the bounding box of the can when present. Which one of the following do you agree with the most?

- We can approach the task as an image classification with a localization problem.
- The two-step system is always a better option compared to an end-to-end solution.
- An end-to-end solution is always superior to a two-step system.
- We can't solve the task as an image classification with a localization problem since all the bounding boxes have the same dimensions.

Expand

Correct
Correct. We can use a network to combine the two tasks similar to that described in the lectures.

3. When building a neural network that inputs a picture of a person's face and outputs N landmarks on the face (assume that the input image contains exactly one face), we need two coordinates for each landmark, thus we need 2N output units. True/False? 1 / 1 point

- (•) True
- ( ) False

Expand

Correct
Correct. Recall that each landmark is a specific position in the face's image, thus we need to specify two coordinates for each landmark.

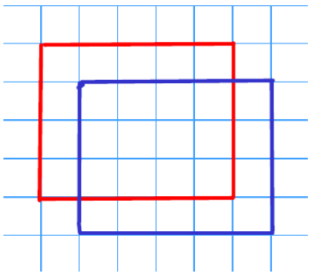
4. When training one of the object detection systems described in the lectures, you need a training set that contains many pictures of the object(s) you wish to detect. However, bounding boxes do not need to be provided in the training set, since the algorithm can learn to detect the objects by itself. 1 / 1 point

- ( ) True
- (•) False

Expand

Correct
Correct, you need bounding boxes in the training set. Your loss function should try to match the predictions for the bounding boxes to the true bounding boxes from the training set.

5. What is the IoU between the red box and the blue box in the following figure? Assume that all the squares have the same measurements. 0 / 1 point



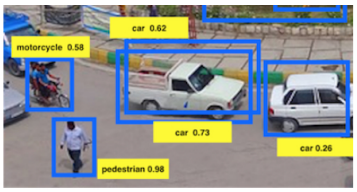
- ( ) $\frac{4}{5}$
- ( ) $\frac{3}{7}$
- ( )

Loading [MathJax]/jax/output/CommonHTML/jax.js

Expand

Incorrect
Recall that IoU is calculated as the quotient of the area of the intersection over the area of the union.

6. Suppose you run non-max suppression on the predicted boxes below. The parameters you use for non-max suppression are that boxes with probability ≤ 0.4 are discarded, and the IoU threshold for deciding if two boxes overlap is 0.5. How many boxes will remain after non-max suppression? 1 / 1 point



- (•) 5
- ( ) 4
- ( ) 7
- ( ) 3
- ( ) 6

Expand

Correct
Correct!

7. Suppose you are using YOLO on a 19x19 grid, on a detection problem with 20 classes, and with 5 anchor boxes. During training, for each image you will need to construct an output volume y as the target value for the neural network; this corresponds to the last layer of the neural network. (y may include some "?", or "don't cares"). What is the dimension of this output volume? **1 / 1 point**

- ( ) 19x19x(25x20)
- ( ) 19x19x(20x25)
- ( ) 19x19x(5x20)
- (•) 19x19x(5x25)

Expand

Correct
Correct, you get a 19x19 grid where each cell encodes information about 5 boxes and each box is defined by a confidence probability (p_c), 4 coordinates (b_x, b_y, b_h, b_w) and classes ($c_1, \ldots, c_{20}$).

8. We are trying to build a system that assigns a value of 1 to each pixel that is part of a tumor from a medical image taken from a patient. **1 / 1 point**

This is a problem of localization? True/False

- (•) False
- ( ) True

Expand

Correct
Correct. This is a problem of semantic segmentation since we need to classify each pixel from the image.

9. Using the concept of Transpose Convolution, fill in the values of **X**, **Y** and **Z** below. **0 / 1 point**

(*padding = 1, stride = 2*)

Input: 2x2

1	2
3	4

Filter: 3x3

1	1	1
0	0	0
-1	-1	-1

Result: 6x6

	0	0	0	X	
	Y	4	2	2	
	0	0	0	0	
	-3	Z	-4	-4	

- X = 0, Y =-1, Z = -7
- X = 0, Y = 2, Z = -1
- X = 0, Y = 2, Z = -7
- X = 0, Y = -1, Z = -4

Expand

Incorrect
To revise the concepts watch the lecture .

10. Suppose your input to a U-Net architecture is h x w x 3, where 3 denotes your number of channels (RGB). What will be the dimension of your output ? 1 / 1 point

- h x w x n, where n = number of of output channels
- h x w x n, where n = number of output classes
- h x w x n, where n = number of filters used in the algorithm
- h x w x n, where n = number of input channels

Expand

Correct